

A closed-form expression for the Sharma-Mittal entropy of exponential families

Frank Nielsen^{1,2} and Richard Nock³

¹ Sony Computer Science Laboratories, Inc. Tokyo, Japan.

² École Polytechnique, CS Dept. LIX, Palaiseau, France.

³ UAG-CEREGMIA, Martinique, France.

E-mail: Frank.Nielsen@acm.org, Richard.Nock@martinique.univ-ag.fr

Abstract. The Sharma-Mittal entropies generalize the celebrated Shannon, Rényi and Tsallis entropies. We report a closed-form formula for the Sharma-Mittal entropies and relative entropies for arbitrary exponential family distributions. We instantiate explicitly the formula for the case of the multivariate Gaussian distributions and discuss on its estimation.

PACS numbers: 03.65.Ta, 03.67.-a

Submitted to: *Journal of Physics A: Mathematical and Theoretical*
Fast Track Communications (revised manuscript, Nov. 2011; accepted Nov. 2011)

The Sharma-Mittal entropy $H_{\alpha,\beta}(p)$ [1, 2] of a probability density[‡] p is defined as

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} \left(\left(\int p(x)^\alpha dx \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \text{ with } \alpha > 0, \alpha \neq 1, \beta \neq 1. \quad (1)$$

This bi-parametric family of entropies tends in limit cases to Rényi entropies $R_\alpha(p) = \frac{1}{1-\alpha} \log \int p(x)^\alpha dx$ (for $\beta \rightarrow 1$), Tsallis entropies $T_\alpha(p) = \frac{1}{1-\alpha} (\int p(x)^\alpha dx - 1)$ (for $\beta \rightarrow \alpha$), and Shannon entropy $H(p) = -\int p(x) \log p(x) dx$ (for both $\alpha, \beta \rightarrow 1$). Sharma-Mittal entropy has previously been studied in the context of multidimensional harmonic oscillator systems [3].

Many usual statistical distributions including the Gaussians and discrete multinomials (that is, normalized histograms) belong to the exponential families [4]. Those exponential families play a major role in the field of thermo-statistics [5], and admit the generic canonical decomposition

$$p_F(x|\theta) = \exp(\langle \theta, t(x) \rangle - F(\theta) + k(x)), \quad (2)$$

where $\langle \cdot, \cdot \rangle$ denote the inner product, F is a strictly convex C^∞ function characterizing the family (called the log-normalizer since $F(\theta) = \log \int e^{\langle \theta, t(x) \rangle + k(x)} dx$), $\theta \in \Theta$ is the natural parameter denoting the member of the family $\mathcal{E}_F = \{p_F(x|\theta) \mid \theta \in \Theta\}$, $t(x)$ is the sufficient statistics, and $k(x)$ is an auxiliary carrier measure [4]. The natural parameter space $\Theta = \{\theta \mid p_F(x; \theta) < \infty\}$ is an open convex set.

For example, the probability density of a multivariate Gaussian $p \sim N(\mu, \Sigma)$ centered at μ with positive-definite covariance matrix Σ is conventionally written as

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|\Sigma|}} \exp - \frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2}, \quad (3)$$

where $|\Sigma| > 0$ denote the determinant of the positive definite matrix. Rewriting the density of Eq. 3 to fit the canonical decomposition of Eq. 2, we get

$$p(x|\mu, \Sigma) = \exp \left(-\frac{1}{2} x^T \Sigma^{-1} x + x^T \Sigma^{-1} \mu - \frac{1}{2} \mu^T \Sigma^{-1} \mu - \frac{1}{2} \log(2\pi)^d |\Sigma| \right). \quad (4)$$

Using the matrix trace cyclic property, we have $-\frac{1}{2} x^T \Sigma^{-1} x = \text{tr}(-\frac{1}{2} x^T \Sigma^{-1} x) = \text{tr}(x x^T \times (-\frac{1}{2} \Sigma^{-1}))$, where tr denote the matrix trace operator. It follows that

$$p(x|\mu, \Sigma) = \exp \left(\left\langle (x, x x^T), \left(\Sigma^{-1} \mu, -\frac{1}{2} \Sigma^{-1} \right) \right\rangle - F(\theta) \right), \quad (5)$$

$$= p(x|\theta). \quad (6)$$

with $\theta = (\Sigma^{-1} \mu, -\frac{1}{2} \Sigma^{-1})$ and $F(\theta) = \frac{1}{2} \log(2\pi)^d |\Sigma| + \frac{1}{2} \mu^T \Sigma^{-1} \mu$ (and $k(x) = 0$). In this decomposition, the natural parameter $\theta = (\Sigma^{-1} \mu, -\frac{1}{2} \Sigma^{-1}) = (v, M)$ consists of two

[‡] For sake of simplicity and without loss of generality, we consider the probability density function p of a continuous random variable $X \sim p$ in this note. For multivariate densities p , the integral notation \int denote the corresponding multi-dimensional integral, so that we write for short $\int p(x) dx = 1$. Our results hold for probability mass functions, and probability measures in general.

parts: a vectorial part v , and a symmetric negative definite matrix part $M \prec 0$. The inner product of $\theta = (v, M)$ and $\theta' = (v', M')$ is defined as $\langle \theta, \theta' \rangle = v^T v' + \text{tr}(M^T M')$. For univariate normal distributions, the natural parameter θ is $(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$. The order of the exponential family is the dimension of its natural parameter space Θ . Normal d -dimensional distributions $N(\mu, \Sigma)$ form an exponential family of order $d + \frac{d(d+1)}{2} = \frac{d(d+3)}{2}$.

We have $M = -\frac{1}{2}\Sigma^{-1}$, that is $|\Sigma^{-1}| = |\Sigma|^{-1} = |-2M|$, and $\mu^T = -\frac{1}{2}v^T M^{-1}$ (since $M^{-1} = -2\Sigma$, $-\frac{1}{2}M^{-1}v = \Sigma v = \mu$ and $M^{-T} = M^{-1}$). It follows that the log-normalizer F expressed using the canonical natural parameters is

$$F(\mu, \Sigma) = \frac{1}{2} \log(2\pi)^d |\Sigma| + \frac{1}{2} \mu^T \Sigma^{-1} \mu, \quad (7)$$

$$F(v, M) = \frac{d}{2} \log 2\pi - \frac{1}{2} \log |-2M| - \frac{1}{4} v^T M^{-1} v. \quad (8)$$

In order to calculate the Sharma-Mittal entropy of Eq. 1, let $M_\alpha(p) = \int p(x)^\alpha dx$ so that

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} \left(M_\alpha(p)^{\frac{1-\beta}{1-\alpha}} - 1 \right). \quad (9)$$

Let us prove that for an arbitrary exponential family $\mathcal{E}_F = \{p_F(x|\theta) \mid \theta \in \Theta\}$

$$M_\alpha(p) = e^{F(\alpha\theta) - \alpha F(\theta)} E_p[e^{(\alpha-1)k(x)}] \quad (10)$$

Proof:

$$M_\alpha(p) = \int e^{\alpha(\langle t(x), \theta \rangle - F(\theta) + k(x))} dx, \quad (11)$$

$$= \int e^{\langle t(x), \alpha\theta \rangle - \alpha F(\theta) + \alpha k(x) + (1-\alpha)k(x) - (1-\alpha)k(x) + F(\alpha\theta) - F(\alpha\theta)} dx, \quad (12)$$

$$= \int e^{F(\alpha\theta) - \alpha F(\theta)} p_F(x; \alpha\theta) e^{(\alpha-1)k(x)} dx, \quad (13)$$

$$= e^{F(\alpha\theta) - \alpha F(\theta)} \int p_F(x; \alpha\theta) e^{(\alpha-1)k(x)} dx, \quad (14)$$

$$= e^{F(\alpha\theta) - \alpha F(\theta)} E_p[e^{(\alpha-1)k(x)}]. \quad (15)$$

Observe that in Eq. 13, we require $\alpha\theta \in \Theta$ for a valid exponential family distribution. This is the case whenever the natural parameter space Θ is a convex cone (e.g., Gaussian case). It follows from Eq. 10 that the Sharma-Mittal entropy of a distribution $p \sim \mathcal{E}_F$ belonging to an exponential family \mathcal{E}_F is

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} \left(e^{\frac{1-\beta}{1-\alpha}(F(\alpha\theta) - \alpha F(\theta))} (E_p[e^{(\alpha-1)k(x)}])^{\frac{1-\beta}{1-\alpha}} - 1 \right). \quad (16)$$

In particular, when the auxiliary carrier measure $k(x) = 0$ [4] (including the above-mentioned multivariate Gaussian family), Eq. 16 becomes a closed-form formula since $E_p[e^{(\alpha-1)k(x)}] = E_p[1] = 1$:

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} \left(e^{(1-\beta) \frac{F(\alpha\theta) - \alpha F(\theta)}{1-\alpha}} - 1 \right), \quad (17)$$

$$= \frac{1}{1-\beta} \left((e^{F(\alpha\theta) - \alpha F(\theta)})^{\frac{1-\beta}{1-\alpha}} - 1 \right). \quad (18)$$

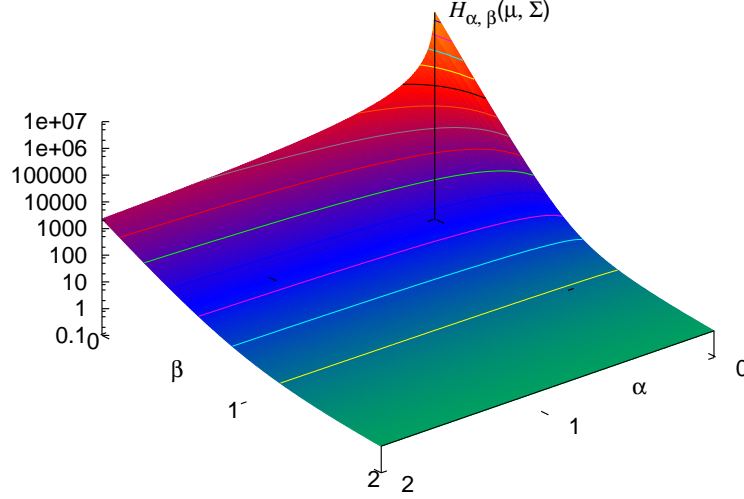


Figure 1. Plot of the Sharma-Mittal entropy (Eq. 22) for the 4×4 covariance matrix $\Sigma = 4I$ (independent of the mean μ), where I denotes the identity matrix.

We derive in limit cases expressions for the Rényi, Tsallis and Shannon entropies of an arbitrary exponential family (with $k(x) = 0$):

$$R_\alpha(p) = \lim_{\beta \rightarrow 1} H_{\alpha, \beta}(p) = \frac{1}{1 - \alpha} (F(\alpha\theta) - \alpha F(\theta)), \quad (19)$$

$$T_\alpha(p) = \lim_{\beta \rightarrow \alpha} H_{\alpha, \beta}(p) = \frac{1}{1 - \alpha} (e^{F(\alpha\theta) - \alpha F(\theta)} - 1), \quad (20)$$

$$H(p) = \lim_{\beta, \alpha \rightarrow 1} H_{\alpha, \beta}(p) = F(\theta) - \langle \theta, \nabla F(\theta) \rangle. \quad (21)$$

Note that the Shannon entropy of a member of an exponential family $p \sim \mathcal{E}_F$ indexed with natural parameter θ can also be rewritten as $H(p) = H(\theta) = -F^*(\eta)$ with $\eta = \nabla F(\theta)$ the dual moment coordinates, and F^* the Legendre C^∞ convex conjugate of F [6].

Let us instantiate the generic formula of Eq. 17 to the case of multivariate Gaussians with mean parameter μ and covariance matrix Σ . We get

$$H_{\alpha, \beta}(\mu, \Sigma) = \frac{1}{1 - \beta} \left(\frac{\left((2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \right)^{1-\beta}}{\alpha^{\frac{d(1-\beta)}{2(1-\alpha)}}} - 1 \right), \quad (22)$$

independent of μ (in 1D, $\Sigma = \sigma^2$ so that $|\Sigma|^{\frac{1}{2}} = \sigma$). Indeed, consider the expression $F(\alpha\theta) - \alpha F(\theta)$ in Eq. 16 for the Gaussian log-normalizer of Eq. 8. Using the fact that $|\alpha A| = \alpha^d |A|$ for d -dimensional matrices A , the term $F(\alpha\theta)$ is

$$F(\alpha\theta) = F(\alpha v, \alpha M), \quad (23)$$

$$= \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\alpha^d| - 2M| - \frac{1}{4} (\alpha v)^T (\alpha M)^{-1} (\alpha v), \quad (24)$$

$$= \frac{d}{2} \log 2\pi - \frac{d}{2} \log \alpha - \frac{1}{2} \log | - 2M| - \frac{\alpha}{4} v^T M^{-1} v. \quad (25)$$

Similarly, we have

$$\alpha F(\theta) = \alpha F(v, M) = \frac{d\alpha}{2} \log 2\pi - \frac{\alpha}{2} \log | - 2M| - \frac{\alpha}{4} v^T M^{-1} v. \quad (26)$$

Thus by subtracting Eq. 26 to Eq. 25, we obtain

$$F(\alpha\theta) - \alpha F(\theta) = \frac{d(1-\alpha)}{2} \log 2\pi - \frac{d}{2} \log \alpha - \frac{1-\alpha}{2} \log | - 2M|. \quad (27)$$

Therefore, we deduce that

$$F(\alpha\theta) - \alpha F(\theta) = \log \left(\frac{(2\pi)^{\frac{d(1-\alpha)}{2}}}{\alpha^{\frac{d}{2}}} | - 2M|^{-\frac{1-\alpha}{2}} \right), \quad (28)$$

$$= \log \left(\frac{(2\pi)^{\frac{d}{2}(1-\alpha)}}{\alpha^{\frac{d}{2}}} |\Sigma|^{\frac{1-\alpha}{2}} \right), \quad (29)$$

hence the result of Eq. 22. For 1D Gaussians with standard deviation $\sigma > 0$, this yields

$$H_{\alpha,\beta}(\mu, \sigma) = \frac{1}{1-\beta} \left(\frac{(\sqrt{2\pi}\sigma)^{1-\beta}}{\alpha^{\frac{(1-\beta)}{2(1-\alpha)}}} - 1 \right). \quad (30)$$

Note that the differential Sharma-Mittal entropy of Gaussians may potentially be negative.

Figure 1 displays the plot of the Sharma-Mittal entropy for a 4×4 covariance matrix set to $4I$, where I denotes the identity matrix.

We also report respectively the Rényi, Tsallis and Shannon entropies for multivariate Gaussians

$$R_\alpha(\mu, \Sigma) = \log \frac{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}{\alpha^{\frac{d}{2(1-\alpha)}}}, \quad (31)$$

$$T_\alpha(\mu, \Sigma) = \frac{1}{1-\alpha} \left(\frac{((2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}})^{1-\alpha}}{\alpha^{\frac{d}{2}}} - 1 \right), \quad (32)$$

$$H(\mu, \Sigma) = \log \sqrt{(2\pi e)^d |\Sigma|} = \frac{1}{2} (d + d \log 2\pi + \log |\Sigma|). \quad (33)$$

Figure 2 displays the plots of the Rényi (Eq. 31) and Tsallis (Eq. 32) entropies for a $d \times d$ -dimensional covariance matrix $\Sigma = \sigma^2 I = 4I$ ($\sigma = 2$).

Information geometry [6] considers the underlying differential geometry induced by a divergence. From the Sharma-Mittal entropy, we can derive the Sharma-Mittal divergence [7] between two distributions $P \sim p$ and $Q \sim q$

$$D_{\alpha,\beta}(p : q) = \frac{1}{\beta - 1} \left(\left(\int p(x)^\alpha q(x)^{1-\alpha} dx \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \forall \alpha > 0, \alpha \neq 1, \beta \neq 1. \quad (34)$$

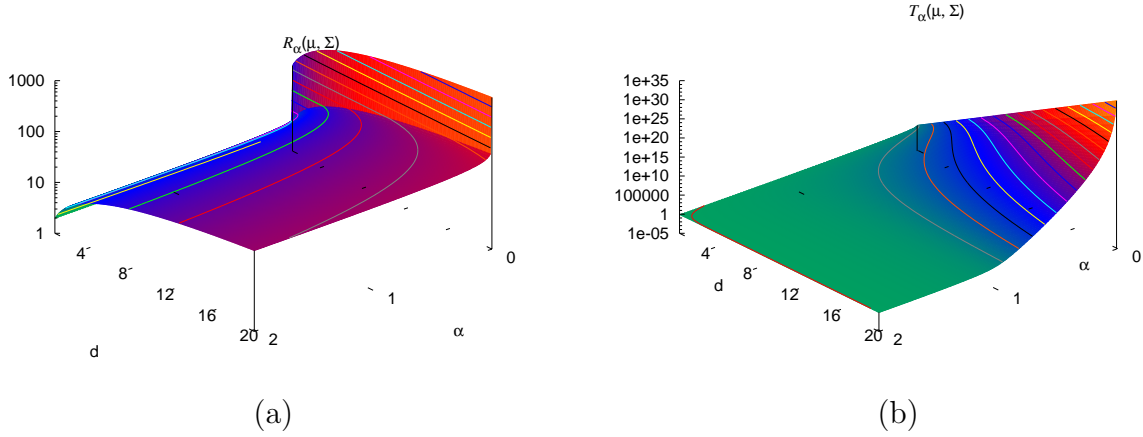


Figure 2. Plot of the (a) Rényi (Eq. 31) and (b) Tsallis (Eq. 32) entropies for covariance matrices set to four times the matrix identity.

Note that $D_{\alpha,\beta}(p : q) = 0$ if and only if $p = q$, since in that case $\int p(x)^\alpha q(x)^{1-\alpha} dx = \int p(x) dx = 1$.

For $\alpha, \beta \rightarrow 1$, the divergence tends to the renowned Kullback-Leibler divergence. Let $C_\alpha(p : q) = \int p(x)^\alpha q(x)^{1-\alpha} dx$ denote the α -divergence [6], related to the Hellinger integral of order α : $H_\alpha(p, q) = 1 - C_\alpha(p, q)$. For $\alpha = \frac{1}{2}$, the similarity measure $C_{\frac{1}{2}}(p : q)$ is symmetric and called the Bhattacharyya coefficient. The Bhattacharyya coefficient is related to the following squared Hellinger distance:

$$H^2(p : q) = \frac{1}{2} \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx = 1 - C_{\frac{1}{2}}(p : q). \quad (35)$$

We rewrite compactly the Sharma-Mittal divergence of Eq. 34 as

$$D_{\alpha,\beta}(p : q) = \frac{1}{\beta - 1} \left(C_\alpha(p, q)^{\frac{1-\beta}{1-\alpha}} - 1 \right). \quad (36)$$

Let us prove that for members $p(x) = p_F(x|\theta)$ and $q = p_F(x|\theta')$ belonging to the *same* exponential family \mathcal{E}_F , we have $C_\alpha(p : q) = e^{-J_{F,\alpha}(\theta:\theta')}$, where

$$J_{F,\alpha}(\theta : \theta') = \alpha F(\theta) + (1 - \alpha) F(\theta') - F(\alpha\theta + (1 - \alpha)\theta') \quad (37)$$

is a Jensen difference divergence [8].

Proof:

$$C_\alpha(p : q) = \int p_F(x|\theta)^\alpha p_F(x|\theta')^{1-\alpha} dx, \quad (38)$$

$$\begin{aligned} C_\alpha(\theta : \theta') &= \int \exp^{\alpha(\langle t(x), \theta \rangle - F(\theta) + k(x))} \\ &\quad \times \exp^{(1-\alpha)(\langle t(x), \theta' \rangle - F(\theta') + k(x))} dx, \end{aligned} \quad (39)$$

$$\begin{aligned} &= \int e^{\langle t(x), \alpha\theta + (1-\alpha)\theta' \rangle} \\ &\quad \times \exp^{-\alpha F(\theta) - (1-\alpha)F(\theta') + k(x)} dx, \end{aligned} \quad (40)$$

$$= \int e^{F(\alpha\theta + (1-\alpha)\theta') - \alpha F(\theta) - (1-\alpha)F(\theta')} p_F(x; \alpha\theta + (1 - \alpha)\theta') dx, \quad (41)$$

$$= e^{-J_{F,\alpha}(\theta:\theta')} \int p_F(x; \alpha\theta + (1-\alpha)\theta') dx, \quad (42)$$

$$= e^{-J_{F,\alpha}(\theta:\theta')} > 0. \quad (43)$$

Observe that for $\alpha \in (0, 1)$, $\alpha\theta + (1-\alpha)\theta' \in \Theta$ since Θ is an open convex set, and therefore the distribution $p_F(x; \alpha\theta + (1-\alpha)\theta')$ is well-defined in Eq. 41.

It follows that the Sharma-Mittal divergence of distributions belonging to the same exponential family (even when $k(x) \neq 0$) is the following closed-form formula

$$D_{\alpha,\beta}(p : q) = \frac{1}{\beta - 1} \left(C_\alpha(\theta_p : \theta_q)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \quad (44)$$

$$= \frac{1}{\beta - 1} \left(e^{-\frac{1-\beta}{1-\alpha} J_{F,\alpha}(\theta_p : \theta_q)} - 1 \right). \quad (45)$$

For multivariate Gaussians, let us explicit the Jensen difference divergence $J_{F,\alpha}$ as the difference of two terms using the (μ, Σ) coordinate system

$$\begin{aligned} \alpha F(\theta) + (1-\alpha)F(\theta') &= \frac{d}{2} \log 2\pi + \frac{1}{2} \log |\Sigma|^\alpha |\Sigma'|^{1-\alpha} \\ &\quad + \frac{\alpha}{2} \mu^T \Sigma^{-1} \mu + \frac{1-\alpha}{2} \mu'^T \Sigma'^{-1} \mu' \end{aligned} \quad (46)$$

and $F(\alpha\theta + (1-\alpha)\theta')$. Let

$$\bar{\theta}_\alpha = \alpha\theta + (1-\alpha)\theta' = (\alpha\Sigma^{-1}\mu + (1-\alpha)\Sigma'^{-1}\mu', -\frac{\alpha}{2}\Sigma^{-1} - \frac{1-\alpha}{2}\Sigma'^{-1}) = (\bar{v}_\alpha, \bar{M}_\alpha). \quad (47)$$

Denote by $\bar{\Sigma}_\alpha = -\frac{1}{2}\bar{M}_\alpha^{-1}$ and $\bar{\mu}_\alpha = \bar{v}_\alpha \bar{\Sigma}_\alpha$ the corresponding parameters. Using Eq. 7, we have

$$F(\alpha\theta + (1-\alpha)\theta') = F(\bar{\mu}_\alpha, \bar{\Sigma}_\alpha) = \frac{1}{2} \log(2\pi)^d |\bar{\Sigma}_\alpha| + \frac{1}{2} \bar{\mu}_\alpha^T \bar{\Sigma}_\alpha^{-1} \bar{\mu}_\alpha. \quad (48)$$

It follows that the Jensen difference divergence between two Gaussian distributions $p \sim N(\mu, \Sigma)$ and $q \sim N(\mu', \Sigma')$ is given by the closed-form formula

$$J_{F,\alpha}(p : q) = \frac{1}{2} \left(\log \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} + \alpha \mu^T \Sigma^{-1} \mu + (1-\alpha) \mu'^T \Sigma'^{-1} \mu' - \bar{\mu}_\alpha^T \bar{\Sigma}_\alpha^{-1} \bar{\mu}_\alpha \right), \quad (49)$$

with

$$\bar{\Sigma}_\alpha = \left(\alpha\Sigma^{-1} + (1-\alpha)\Sigma'^{-1} \right)^{-1}, \quad (50)$$

$$\bar{\mu}_\alpha = \bar{\Sigma}_\alpha \bar{v}_\alpha = \bar{\Sigma} \left(\alpha\Sigma^{-1}\mu + (1-\alpha)\Sigma'^{-1}\mu' \right) \quad (51)$$

Letting $\Delta\mu = \mu' - \mu$, Eq. 49 can further be rewritten compactly as

$$J_{F,\alpha}(p : q) = \frac{1}{2} \left(\log \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} + \alpha(1-\alpha) \Delta\mu^T \bar{\Sigma}_\alpha^{-1} \Delta\mu \right). \quad (52)$$

Thus we obtain a closed-form formula for the Sharma-Mittal divergence of multivariate Gaussians generalizing the Rényi α -divergences, formerly reported in [9]:

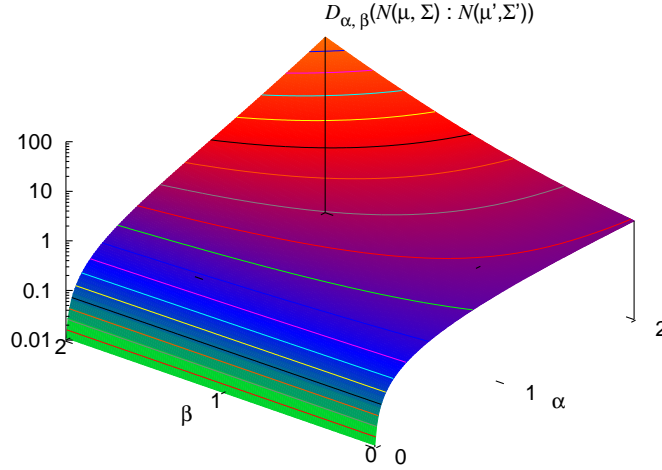


Figure 3. Plot of the Sharma-Mittal divergence $D_{\alpha, \beta}$ (Eq. 54) for univariate normal distributions with respective standard deviation $\sigma = \sqrt{2}$ and $\sigma' = 2$, and mean difference $\mu - \mu' = 4$.

$$D_{\alpha, \beta}(N(\mu, \Sigma) : N(\mu', \Sigma')) = \frac{1}{\beta - 1} \left(e^{-\frac{1-\beta}{2(1-\alpha)} \left(\log \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} + \alpha(1-\alpha) \Delta \mu^T \bar{\Sigma}_\alpha^{-1} \Delta \mu \right)} - 1 \right), \quad (53)$$

$$= \frac{1}{\beta - 1} \left(\left(\frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} \right)^{-\frac{1-\beta}{2(1-\alpha)}} \exp \left(-\frac{\alpha(1-\beta)}{2} \Delta \mu^T \bar{\Sigma}_\alpha^{-1} \Delta \mu \right) - 1 \right). \quad (54)$$

Figure 3 shows the plot of the Sharma-Mittal divergence for univariate normal distributions with respective standard deviation $\sigma = \sqrt{2}$ and $\sigma' = 2$, and mean difference $\mu - \mu' = 4$ in Eq. 54. In practice, for numerical stability, we prefer to compute the divergence by first computing the Jensen difference divergence of Eq. 52, and then applying generic formula of Eq. 45.

The underlying distribution is usually not explicitly given so that we need to first estimate the distribution or related quantities like its entropy [10]. Leonenko et al. [11] proposed a method to estimate entropies using the k -nearest neighbor graph (k -NN) of an independently and identically distributed sample set x_1, \dots, x_n . However, their method suffers from the curse of dimensionality of computing k -NN graphs and falls short when dealing with moderate dimensions. For exponential families, we can estimate the natural parameter of an exponential family using the maximum likelihood estimator (MLE) that admits the unique global optimum [4, 5] $\hat{\theta}$ such that

$$\nabla F(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n t(x_i). \quad (55)$$

For multivariate Gaussians, from the sufficient statistic $t(x) = (x, x^T x)$ we deduce that $\nabla F(\hat{\theta}) = (\hat{\mu}, \hat{\Sigma} + \hat{\mu} \hat{\mu}^T) = (\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n x_i^T x_i)$.

It follows a simple and fast scheme to estimate the Sharma-Mittal entropy (or divergence) from n observations sampled identically and independently from an exponential family distribution: Estimate the natural parameter using Eq. 55 and apply formula of Eq. 17.

To conclude, let us note that *any* arbitrary smooth density can be approximated by an exponential family of order depending on the approximation precision [12, 13] (enforcing no extra auxiliary carrier measure: That is, with $k(x) = 0$). Thus we can approximate the Sharma-Mittal entropy of an arbitrary probability density [13] by approximating it to a close exponential family, and then applying the closed-form formula Eq. 17. We believe that Eq. 17, Eq. 22, Eq. 45 (numerically stable) and 54 will prove useful when experimenting for suitable parameters (α, β) in various statistical signal tasks [9].

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions. Frank Nielsen (5793b870) would like to thank Dr. Kitano and Dr. Tokoro for their support. Additional material is available on-line at <http://www.informationgeometry.org/SharmaMittal/>

References

- [1] B. D. Sharma and D. P. Mittal. New nonadditive measures of inaccuracy. *J. Math. Sci.*, 10:122–133, 1975.
- [2] E. Akturk, G. B. Bagci, and R. Sever. Is Sharma-Mittal entropy really a step beyond Tsallis and Rényi entropies? *arXiv*, cond-mat/0703277, 2007.
- [3] O. Üzengi Aktürk and E. Aktürk and M. Tomak. Can Sobolev inequality be written for Sharma-Mittal entropy? *Int. J. Theo. Phys.*, 12(47):3310–3320, 2008.
- [4] L. D. Brown. *Fundamentals of statistical exponential families: with applications in statistical decision theory* Inst. Math. Stat., ISBN:0-940-60010-2, 1986.
- [5] J. Naudts. *Generalised thermostatics*. Springer, 2011.
- [6] S.-i. Amari and H. Nagaoka. *Methods of information geometry*. Oxford University Press, 2000.
- [7] M. Masi. A step beyond Tsallis and Rényi entropies. *Phys. Let. A*, 338(3-5):3–5, 2005.
- [8] J. Burbea and C. R. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theo.*, IT-28(3):489–495, 1982.
- [9] A. O. Hero, B. Ma, O. Michel, and J. D. Gorman. Applications of entropic spanning graphs. *IEEE Sig. Proc. Mag.*, (5), 2002.
- [10] B. Póczos and J. Schneider. On the estimation of α -divergences. In *Proc. 14th Int. Conf. AI and Stat.*, 2011.
- [11] N. Leonenko, L. Pronzato, and V. Savani. A class of Rényi information estimators for multidimensional densities. *Ann. Stat.*, 36(5):2153–2182, 2008.
- [12] L. Cobb, P. Koppstein, and N. H. Chen. Estimation and moment recursion relations for multimodal distributions of the exponential family. *J. Am. Stat. Assoc.*, 78(381):124–130, 1983.
- [13] D. C. Brody. A note on exponential families of distributions *J. Phys. A: Math. and Theo.*, 78(381):124–130, 2007.